Secure Space-Time Communication

Alfred O. Hero

April 24, 2001

Submitted to *IEEE Transactions on Information Theory*, April 2001-

Abstract

Network security is important for information protection in open, secure or covert wireless communications. One such requirement is to achieve high rate communications between clients in the network while hiding information about the transmitted symbols, signal activity, or other sensitive data from an unintended receiver, e.g. an eavesdropper. For wireless links the single user capacity advantages of deployment of multiple antennas at the transmitter is well known. One of the principal conclusions of this paper is that proper exploitation of space-time diversity at the transmitter can also enhance information security and information hiding capabilities. In particular, we show that significant gains are achievable when the transmitter and the client receiver are both informed about their channel while the transmitter and eavesdropper receiver are uniformed about their channel. More generally, we compare capacity limits for both informed and uninformed transmitter and informed receiver scenarios subject to low probability of intercept (LPI) and low probability of detection (LPD) constraints. For several general cases we can characterize the LPI- and LPD-optimal transmitted source distributions and compare them to the standard optimal source distribution satisfying a power constraint. We assume the widespread quasi-static flat Rayleigh fading channel model for the transmitter-receiver pairs. This paper is a step towards answering the fundamental question: what are the qualitative and quantitative differences between the information carrying capabilities of open space-time channels versus secure space-time channels?

Keywords: space-time coding, space-time channel capacity, covert channels, wireless eavesdroppers, information hiding, perfect secrecy, Chernoff exponents, flat Rayleigh fading models.

Corresponding author:

Professor Alfred O. Hero III
Room 4229, Dept. of EECS
University of Michigan
1301 Beal Avenue
Ann Arbor, MI 48109-2122.
Tel: (734) 763-0564.
Fax: (734) 763-8041.
e-mail: hero@eecs.umich.edu.

[1]A. Hero is with the Dept. of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, MI 48109-2122 `hero@eecs.umich.edu`. This work was partially supported by the Army Research Office under grant ARO DAAH04-96-1-0337. Portions of this work were presented at the National Radio Science Meeting of the International Union of Radio Sciences (URSI), Boulder CO, Jan. 2001.

1 Introduction

Researchers in commercial wireless have primarily focussed on quality of service (QoS) expressed in terms of deliverable information rates, channel capacity and outage capacity, throughput and delay. While these are relevant quality measures there is increasing interest in network security both for assurance of data privacy, reliable user authentication, and protection of information from malicious eavesdroppers intent on discovering network vulnerabilities. The starting point of this paper is that a well-designed secure link should have low probability of intercept (LPI) and low probability of detection (LPD) with respect to an unauthorized eavesdropper. An important question which has motivated the work reported in this paper is: what is the fundamental impact of implementing link-level security measures on information rates and channel capacity? This paper describes a theoretical framework for answering such questions which is based on analyzing the fundamental impact on capacity imposed by different classes of link security constraints.

One cannot hope to ensure security without some cooperation of the transmitter and receiver (client) to put an eavesdropper at a relative disadvantage [30]. One of the most most common forms of cooperation is the use of a cipher [20] to encrypt each data stream transmitted which can only be deciphered at the client receiver using a private shared key. We refer to this method as temporal (single-channel) data encryption, an example of which is the US National Data Encryption Standard (DES) for symmetric data encryption/decryption. Use of temporal encryption is a very flexible measure for preventing unauthorized *interception* of private messages which can be applied to any message sequence without considering the physical layer of the network. Another common form of cooperation is for the transmitter and receiver to adopt information hiding measures [26] to prevent unauthorized *detection* of any signaling activity which could be used, for example, for geolocation of the transmitter. Information hiding is a form of covert encryption which encodes private messages in a background signal or noise process in such a way that the presence of the messages is hidden from those without access to the private key. A well known example is spread spectrum modulation for wireless channels which hides the spectral signature of the signal in the broadband noise background using a pseudo-random convolution sequence as a private key. Another example is watermarking where a owner-identifying watermark is hidden in an image or video signal [3]. The results of this paper can be applied to watermarking of space-time signals. The thesis of this paper is that additional security against detection or interception can be achieved by space-time coding over multiple antennas (or acoustic transducers) at transmitter and receiver. In particular, when such information is available to the transmitter, one can design the spatio-temporal modulation/demodulation to exploit known propagation

and interference characteristics of the channel available to the client but not to the eavesdropper. For the memoryless channels considered here, this corresponds to *spatial* (multi-channel) encryption and information hiding where the shared channel information plays the role of a shared private key that can be used to unlock the message.

It is useful to place the contributions of this paper in the background of previous work. Shannon introduced the information theoretic framework for studying secrecy in communications [30]. As secrecy involves at least three terminals, the tranmsitter, the client and the eavesdropper, the study of achievable information rates for secure communications is a branch of multi-terminal (more than two) information theory [5]. Wyner and co-workers [37, 25] developed the concept of the wire-tap channel for wired links and assessed the impact of secrecy on achievable information rate pairs. This work was extended by Csiszár and Körner to more general broadcast channels in [4]. The possibility of enhancing secrecy by incorporating common knowledge of channel impulse response into the data encryption was identified and exploited by Hassan and co-workers [14] and was applied to single antenna mobile radio links in [9] and [19]. This paper takes a different point of view from previous work in that we evaluate the fundamental impact of transmission secrecy (LPI and LPD) on two-terminal channel capacity in the setting of multi-antenna spatio-temporal quasi-static Rayleigh fading channels.

We investigate the client's channel capacity and the capacity-achieving transmission strategy under LPI and LPD constraints. For this we consider the multiple-input multiple output (MIMO) case where both client and eavesdropper access space-time Rayleigh fading channels. For the discrete time and space channels considered here, the transmitted signals are complex valued $T \times M$ matrices whose rows span T time samples and whose columns span M space samples equal to the number of transmit antennas. We show that when both the transmitter and client receiver know the channel they can exploit this knowledge to achieve improved LPI and LPD beyond those achievable by single-channel systems. Such exploitation is possible to a lesser extent when the transmitter does not know the client's channel. Following the terminology used in [2] we say that the client and/or transmitter are *informed* when they know their channel propagation coefficients, while we say that the eavesdropper's link is *uniformed*, i.e. neither transmitter nor eavesdropper know their propagation coefficients.

The LPI constraint can be imposed by constraining the eavesdropper's channel capacity, cutoff rate, or decoding error probability. For example, when the eavesdropper's channel capacity is significantly lower than the client's capacity the converse to Shannon's channel coding theorem implies that, by setting his rate

between the capacities of the client and the eavesdropper, the transmitter can deprive the eavesdropper of arbitrarily low probability of decoding error while reliably communicating to the client. Here we show that when the eavesdropper is uninformed about his channel the transmitter can enforce zero information rate to the eavesdropper while delivering positive information rate to the client. This LPI condition is equivalent to the *perfect secrecy* condition in cryptography [30]. We derive integral expressions for the perfect-secrecy capacity for the informed receiver and for certain cases characterize the optimal signaling distributions which achieve it. The LPD constraint is imposed by constraining the eavesdropper's probability of correctly detecting the presence of any signaling activity by the transmitter. This is closely related to the *steganography* problem [26]. We make conservative assumptions on the information possessed by the eavesdropper, e.g. the eavesdropper knows only the transmitted signal distribution and the received signal-to-noise ratio (SNR). To obtain tractable expressions for the LPD-constrained capacity we will rely on Chernoff error exponents, large eavesdropper standoff assumptions, and Edgeworth expansions of the eavesdropper's probability densities. The Chernoff exponent defines the asymptotic rate of decrease of the probability of detection error as the block-length of the code goes to infinity. This exponent will be used to define appropriate LPD constraints on transmitted signals.

Most of the results presented here apply to the case where the eavesdropper is at large standoff from the clients link. This implies that the eavesdropper has both low received SNR and approximately Gaussian multi-user interference and noise statistics. In addition, while many of these results can be generalized, we assume that both the client and the eavesdropper access the transmitted energy through distinct mutually-independent quasi-static Rayleigh fading channels [1]. As mentioned above, exploiting channel information known to transmitter and client's receiver can be viewed as a form of spatial encryption where the shared private key is the set of channel propagation coefficients. As a practical matter, a transmitter and receiver informed link requires that private and possibly encrypted training sequences be transmitted to the client and the subsequent channel estimates be transmitted back to the transmitter through some feedback mechanism. On the other hand, a receiver-only informed link requires training but no feedback. In both cases the effect of channel estimation errors and time delays may be significant. While we do investigate the effect of erroneous channel information at the transmitter, we do not focus on the broader channel estimation issues in this paper.

We present the following results in this paper:

1. Under the aforementioned space-time Rayleigh channel informed/uninformed dichotomy it is possible

for the transmitter to communicate reliably to the client while depriving the eavesdropper of any transmitted information whatsoever. Thus the transmitter attains perfect secrecy as defined by Shannon [29]. This can be accomplished by restricting the space-time modulation to a class of complex transmitted matrices whose spatial inner product is equal to a constant $T \times T$ matrix. Two examples of such perfect-secrecy constellations are square unitary space-time codes and quaternion space-time codes [17, 16, 31]. The channel capacity when restricted to these signals is herein called the *perfect-secrecy capacity* for which we give integral forms for the case of an informed transmitter and receiver.

2. When the eavesdropper knows both the signal and his channel exactly, constraining the eavesdropper's Chernoff exponent is equivalent to constraining the mean power over the transmitter antennas, which we call a mean-power constraint. Thus we conclude that in this case no additional countermeasures beyond minimizing average transmitter power are required to enhance security of the client's link.

3. When the channel is unknown but the signal is known to the eavesdropper constraining the Chernoff exponent is equivalent at low SNR to constraining the trace of the fourth moment of the signal matrix.

4. When both channel and signal are unknown (but the signal distribution is known) to the eavesdropper the Chernoff exponent reduces to the sum of two terms: a function of the determinant of the spatio-temporal receiver covariance matrix and a tensor product of the receiver kurtosis and the signal covariance. The kurtosis is defined as the expectation of a four-fold product of the spatio-temporal signal amplitudes. The kurtosis tensor product is non-negative and equal to zero when the received signal is complex Gaussian. As the channel is Gaussian zero kurtosis is only possible when the transmitted signal is non-random. When the kurtosis tensor product increases from zero, as occurs, for example, when elements of the received signal matrices obey an increasingly heavy tailed (super-Gaussian) distribution, the eavesdropper's detection performance degrades. This result is reminiscent of the well known negative kurtosis condition under which blind equalization is possible for an unknown single input single output (SISO) channel with memory [28, 32].

5. Under the scenario where channel and signal are unknown to the eavesdropper, at low SNR the constraint on the Chernoff exponent reduces to a constraint on the trace of the square of the transmitted spatio-temporal signal covariance matrix, which we call the mean-squared-power constraint. Unlike the standard mean-power-constraint this constraint penalizes large spatial power variation of the transmitted signals.

6. For informed transmitter and client receiver operating under the mean-squared power constraint the capacity of the client's link is attained by a Gaussian signaling strategy, called the LPD-optimal strategy. In this signaling strategy the transmitted energy is distributed more evenly over the modes of the channel as compared to the water-pouring solution, called the power-optimal strategy, which is optimal under the standard mean-power constraint.

7. For uninformed transmitter but informed client receiver operating under the mean-squared-power constraint both the capacity and the capacity attaining signaling strategy are of identical form to the standard power-optimal capacity obtained under a mean-power constraint. In this case no additional countermeasures are required to enhance security of the client's link against eavesdropping.

8. The LPD-optimal and power-optimal signaling strategies achieve different information transmission rates for equal signal power or for equal LPD performance as measured by the Chernoff exponent. For fixed Chernoff exponent the power-optimal signal achieves lower information rate than the LPD-optimal signal and conversely. We investigate the relative advantages of power-optimal and LPD-optimal signaling as a function of spatial diversity at the transmitter and received SNR. In particular, while LPD-optimal signaling has no advantage over power-optimal signaling for a single transmit antenna (no diversity), it is shown that almost a factor of two information rate advantage is achievable at low SNR with 16 transmit antennas.

We provide a brief outline of the paper. In Section 2 the Rayleigh fading measurement model is introduced. In Section 3 an integral expression is given for the perfect-secrecy capacity. In Section 4 we give Chernoff error exponents for detection error probability for differing levels of channel and signal information available to the eavesdropper. In Section 5 we provide numerical comparisons illustrating the loss in capacity due to adoption of the LPD strategy.

2 Background

An M-element transmitter antenna array transmits a $T \times M$ signal matrix S over a time interval of T time samples, called the coherent fade sampling interval (Fig. 1). Let X^i denote the signal received by the client over channel H_{TR} and Y^i the signal received by the eavesdropper over a channel H_{TE} (Fig. 2). For notational simplicity, throughout this paper superscripts and subscripts will be used interchangeably when no confusion ensues. We will assume that the two receivers have N_R and N_E receive antennas, respectively.

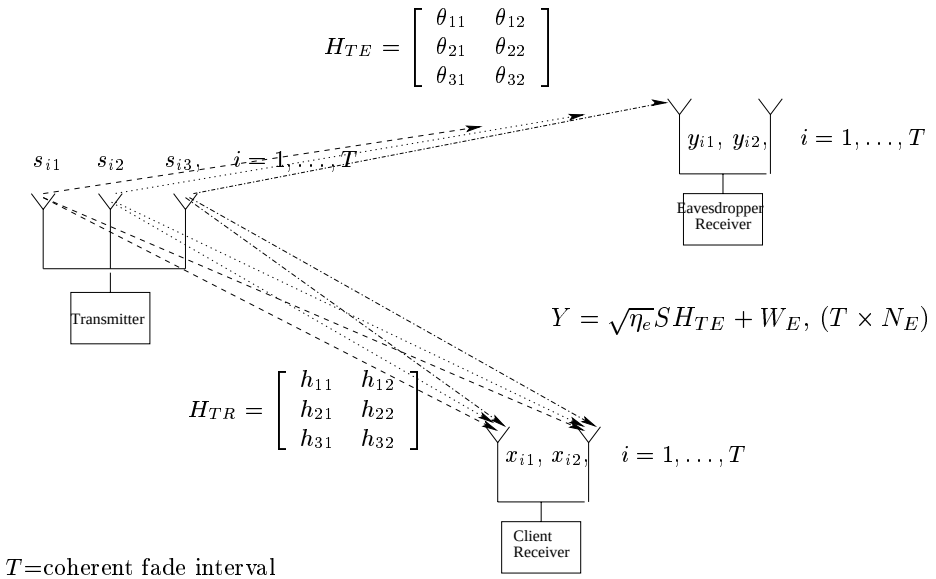

Figure 1: Secure Space-Time Link ($M = 3$, $N_R = N_E = 2$)

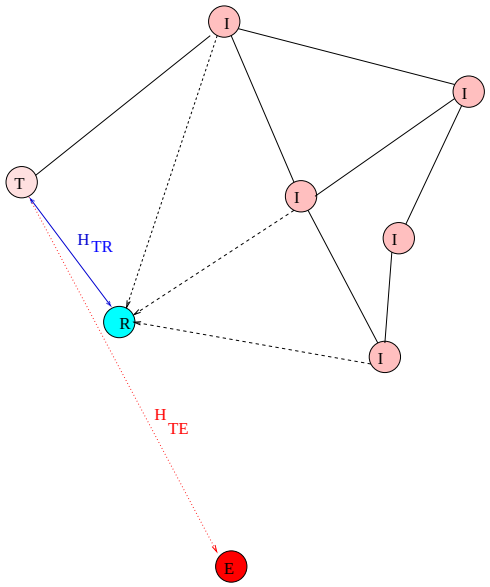

Figure 2: *A link in a wireless network between a transmitter (T) and a receiver (R) who have cooperated to learn their channel H_{TR}. Eavesdropper E attempts to detect a message (known signal), detect signaling activity (known modulation), or intercept data transported by the link H_{TR} without knowing the channel H_{TE}. The eavesdropper and the client receiver must generally perform these tasks in the presence of multi-user interference. In this paper, we assume Gaussian interferers with known spatial covariances.*

Similarly to much previously published research on space time coding [33, 7, 15, 21, 34, 10, 12, 13], we will assume the multi-channel quasi-additive Rayleigh fading models for the received signals. Over L independent frames of T time samples each the models are

$$\begin{aligned} X^i &= \sqrt{\eta_r} S^i H_{TR}^i + W_R^i, \quad i = 1, \ldots, L, \\ Y^i &= \sqrt{\eta_e} S^i H_{TE}^i + W_E^i, \quad i = 1, \ldots, L, \end{aligned} \tag{1}$$

where $S^i \in \mathcal{S}$ is the i-th transmitted signal, $\eta_r = \rho_r/M$, $\eta_e = \rho_e/M$, are the normalized signal-to-noise ratios (SNR) with $\rho_c, \rho_e > 0$ the expected SNR's at each receiver per transmit antenna, H_{TR}^i and H_{TE}^i are mutually independent $M \times N_R$ and $M \times N_E$ matrices of complex channel coefficients, and W_R^i and W_E^i are mutually uncorrelated $T \times N_R$ and $T \times N_E$ matrices of complex circularly symmetric Gaussian noises. Note that we are assuming that H_{TR} and H_{TE} have coherent fade intervals of identical duration T. When multi-user interference is present we can account for it in our capacity calculations by assuming the worst case Gaussian interference scenario. The theory developed here applies to the case where the interferers have known covariances Q_R and Q_E; specifically, Q_R is known to transmitter and client and Q_E is known to transmitter and eavesdropper. In this case the client and eavesdropper models can be reduced to the white noise models (1) by suitable prewhitening at the receivers. Extension of the theory developed in the sequel to unknown Q_R and Q_E is a difficult open problem which we do not consider here. We denote by common notation N, η the quantities N_E, N_R and η_e, η_r when no risk of confusion ensues. The quasi-static Rayleigh flat fading model corresponds to taking the $LN(T+M)$ elements of the matrices $\{H^i\}_{i=1}^L$ and $\{W^i\}_{i=1}^L$ to be i.i.d. complex zero mean (circularly symmetric) Gaussian random variables with unit variance.

Let $\mathbf{Z} = [Z^1, \ldots, Z^L]$ and $\mathbf{S} = [S^1, \ldots, S^L]$ denote the sequence of L measurement and signal matrices, respectively, and H^i the channel matrix of either the client or the eavesdropper over the i-th frame. Under the assumption that the channel matrices are independent over each coherent fade interval, indexed by i, the joint conditional probability density of the observations factors into a product of marginals

$$p(\mathbf{Z}|\mathbf{S}) = \prod_{i=1}^{L} p(Z^i|S^i),$$

where, if the channel $H = H^i$ is known to the receiver

$$p(Z|S) = p(Z|S,H) = \frac{\exp\left(-\mathrm{tr}\{[Z-\eta SH][Z-\eta SH]^\dagger\}\right)}{\pi^{TN}}, \tag{2}$$

while if the channel is unknown to the receiver

$$p(Z|S) = \frac{\exp\left(-\mathrm{tr}\{[I_T + \eta SS^\dagger]^{-1} Z(Z)^\dagger\right)}{\pi^{TN}\,|I_T + \eta SS^\dagger|}, \tag{3}$$

where I_T is the $T \times T$ identity matrix, and $|A| = |\det(A)|$ denotes the magnitude determinant of square matrix A.

2.1 Mean-Power Constrained Capacity

Following the standard random block coding construction of channel capacity, $\mathbf{S} = [S^1, \ldots, S^L]$ is interpreted as a block code consisting of statistically independent symbols drawn from a source distribution $dP(S)$. Define $\vec{S}$ the concatenation of the T rows of one of these symbols, denoted as $T \times M$ matrix S, into a TM-element row vector. The covariance $\mathrm{cov}(S)$ of S is defined as the Hermitian symmetric $TM \times TM$ matrix $\mathrm{cov}(\vec{S}) = E[\vec{S}^H \vec{S}] - E[\vec{S}^H]E[\vec{S}]$. Let P_{avg} be a specified positive constant. For an informed link where both transmitter and receiver know the channel, the channel capacity under the mean transmitted power constraint

$$\mathrm{tr}\{\mathrm{cov}(S)\}/(TM) \leq P_{\mathrm{avg}}, \tag{4}$$

was derived in [33, 34] as

$$\begin{aligned} C_{\mathrm{pow}}^{TR} &= E\left[\max_{P(S)} \log P(X|S,H)/P(X|H)\right] \\ &= TE\left[\ln\left|I_N + \eta_r H^\dagger \Sigma_{\mathrm{pow}} H\right|\right] \\ &= T\sum_i E\left[(\log \mu\lambda_i)^+\right] \end{aligned} \tag{5}$$

where λ_i are the eigenvalues of $\eta_r HH^\dagger$. The capacity is attained by a zero mean circularly symmetric Gaussian source S with covariance $I_T \bigotimes \Sigma_{\mathrm{pow}}$ where $\Sigma_{\mathrm{pow}} = UDU^\dagger$, U are the (right) eigenvectors of $HH^\dagger$, $D = \mathrm{diag}(\sigma_i)$, σ_i are given by water-filling

$$\sigma_i = (\mu - 1/\lambda_i)^+, \qquad i = 1, \ldots, M \tag{6}$$

and $\mu > 0$ is a parameter such that $M^{-1}\sum_{i=1}^M \sigma_i = P_{\mathrm{avg}}$. In the sequel we will call C_{pow}^{TR} the T/R-informed power-capacity and the informed capacity-achieving spatial signal covariance Σ_{pow} will be called the T/R-informed power-optimal signal covariance. Note that the capacity achieving signal matrix has i.i.d. Gaussian rows each having (spatial) covariance Σ_{pow} whose eigenvectors are the modes (columns of U) of H. Note also that the water-filling strategy allocates transmitter energy only to those channel modes which have the highest associated SNR. it can be shown that the optimal receiver applies a beamformer which is matched to the channel H prior to MAP decoding.

When only the receiver has information about the channel, the transmitter cannot exploit the highest SNR modes and the average power constrained channel capacity takes the form [33]

$$C_{\text{pow}}^{R} = \max_{P(S)} E\left[\log P(X|S,H)/P(X|H)\right]$$

We call this capacity the R-informed power-capacity. The capacity achieving source is a $T \times M$ matrix with i.i.d. zero mean circularly symmetric Gaussian elements having identical variances equal to P_{avg}.

For an uninformed link where neither transmitter and receiver know the channel, the channel capacity under an average transmitted power constraint was first investigated in [21].

$$C = \max_{P(S)} E\left[\log P(X|S)/P(X)\right].$$

While approximations have been investigated [21, 38] no closed form expression exists for either the capacity or the capacity achieving source. However, it was shown in [21] that the capacity achieving source has the abstract form

$$S = V\Lambda$$

where V is an isotropically distributed $T \times T$ matrix and Λ is an independent non-negative $T \times M$ diagonal matrix.

3 Low Probability of Intercept: the Perfect-Secrecy Capacity

Here we focus on the LPI strategy of designing transmitter signaling to zero out the channel information rate available to the eavesdropper while maintaining high information rate communication to the client. We motivate this section by considering cutoff rates.

3.1 Motivation: Channel cut-off rate

The channel cut-off rate R_o is a lower bound on the Shannon channel capacity C. Cut-off rate analysis has frequently been adopted to establish practical coding limits [35, 8] as the cut-off rate specifies the highest information rate beyond which sequential decoding becomes impractical [27, 36] and as it is frequently simpler to calculate than channel capacity. The cutoff rate for an uninformed link with quasi-static Rayleigh channel was derived in [13, 12].

3.2 Single-Link Cutoff Rates

For a space-time channel H the cutoff rate has the general expression [13]:

$$R_o \quad = \quad \max_{P_S} -\ln \int\int_{S_1,S_2 \in \mathbb{C}^{T\times M}} dP_S(S_1) dP_S(S_2) \quad e^{-N D(S_1\|S_2)}$$

where the maximization is over suitably constrained source distributions dP_S and $D(S_1\|S_2)$ is a signal dissimilarity measure between pairs of transmitted signals S_1 and S_2. The cutoff rate increases as dissimilarity between pairs of signals increases, i.e. as the average of $\exp(-ND(S_1\|S_2))$ increases. Thus D is directly related to the information transport and decoding limitations imposed by a particular channel.

The following expressions are easily derived for N receivers and received SNR η

1. Transmitter/receiver informed cutoff rate: H known to both T/R

$$D(S_1\|S_2) = \frac{\eta}{4} \operatorname{tr}\left(H^\dagger (S_1 - S_2)^\dagger (S_1 - S_2) H\right)$$

2. Receiver informed cutoff rate: H known to R only

$$D(S_1\|S_2) = \ln \left| I_M + \frac{\eta}{4} (S_1 - S_2)^\dagger (S_1 - S_2) \right|$$

3. Uninformed cutoff rate: H unknown to either T/R [13]

$$D(S_1\|S_2) = \ln \frac{\left| I_T + \frac{\eta}{2}(S_1 S_1^\dagger + S_2 S_2^\dagger) \right|}{\sqrt{\left| I_T + \eta S_1 S_1^\dagger \right| \left| I_T + \eta S_2 S_2^\dagger \right|}}$$

Note that in the T/R-informed case the channel cutoff rate depends on the dissimilarity of the signal pair *after they are received*, i.e. the difference squared between $S_1 H$ and $S_2 H$, while in the R-informed case the cutoff rate depends on the difference squared between the pair of *transmitted* signals. On the other hand, in the uninformed case the cutoff rate depends on the difference between the determinant of the arithmetic mean (numerator) and the geometric mean (denominator) of the conditional received covariances $\operatorname{cov}(X|S_1) = I_T + \eta S_1 S_1^\dagger$ and $\operatorname{cov}(X|S_2) = I_T + \eta S_2 S_2^\dagger$. Thus only temporal information can be used to distinguish between different signals. A user or eavesdropper on an uninformed channel cannot use any spatial information to help decode the symbols since this spatial information is completely unknown to him.

When the source S has constant spatial inner product $SS^\dagger$ the uninformed receiver's absolute blindness to all spatial information can also be deduced directly from the form of the receiver's likelihood function

$l(S) = p(Z|S)$, given in (3), as this function is constant. This also implies that the channel capacity will be equal to zero and the minimum probability of decoding error will be equal to one if a source with constant $SS^\dagger$ is transmitted over a uninformed Rayleigh fading link.

3.3 Perfect-Secrecy Signaling

We conclude that if the eavesdropper has an uninformed channel, his information rate can be reduced to zero if the transmitter adopts a signaling strategy which uses a constellation $\mathcal{S} = \{S\}$ having constant spatial inner product:

$$SS^\dagger = A, \tag{7}$$

where A is a prespecified non-random $T \times T$ matrix. When A is diagonal, many known signal constellations $\{S_i\}_i$ satisfy this *perfect-secrecy* property.

- Doubly unitary codes ($T \geq M$)

$$S_i^\dagger S_i = I_M, \quad S_i S_i^\dagger = \begin{bmatrix} I_M & O \\ O & O \end{bmatrix}$$

  Some instances of such codes are

  - Square unitary codes ($T = M$) [31]: $S_i S_i^\dagger = S_i^\dagger S_i = I_M$
  - Space time QPSK: Quaternion codes [18]: ($T = M = 2$):

$$\mathcal{S} = \left\{ \pm \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \pm \begin{bmatrix} j & 0 \\ 0 & -j \end{bmatrix}, \pm \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \pm \begin{bmatrix} 0 & j \\ j & 0 \end{bmatrix} \right\}$$

- Constant spatial modulus (CM) codes ($T = 1$):

$$S_i = \underline{S}_i^T = [S_{i1}, \cdots, S_{iM}]$$

$$\mathrm{tr}\{S_i S_i^\dagger\} = \|\underline{S}_i\|^2 = 1$$

3.4 Perfect-Secrecy Capacity

Of obvious interest is the channel capacity of a T/R-informed link with signaling limited to the class of signals satisfying the constant spatial inner product condition (7), which we define as the T/R-informed *perfect-secrecy capacity*. In Appendix A we give an integral expression for this capacity. The capacity

achieving source density dP^* satisfies an equalization condition that says in essence that the optimal source should make the instantaneous per-symbol mutual information independent of the particular transmitted symbol S. In the special case that $N_R \geq M$ and the eigenvalues of $HH^\dagger$ are identical, we show that the optimal source distribution is the uniform codeword distribution $dP(S) = dP^*(S) = 1/\mathrm{vol}(\mathcal{S})$ supported on $\mathcal{S}$ where $\mathcal{S} = \{S : SS^\dagger = A\}$ is the perfect secrecy constraint set. When $A = I_T$ and $T \leq M$ this set is the hypersphere $\mathrm{SO}(\mathbb{C}^{TM})$ which is also known as the Stiefel manifold. This source distribution is similar to the isotropically random unitary source introduced in [21] and suggests that if a tesselation of the hypersphere is possible the equispaced constellation constructed on each lattice point in the tessellation might be close to optimal.

An integral expression for the R-informed *perfect-secrecy capacity* is not available. However, we make the following conjecture: for $N_R \geq M$ and the case of an uninformed transmitter but informed receiver a uniform dP_S is close to optimal for any H for large T and large SNR. The intuition is that in this limiting regime the transmitter knows that the receiver can accurately estimate the channel and diagonalize it, thereby converting the channel to $HH^\dagger = I$ for which the uniform source distribution dP^* is optimal. Recent techniques such as that of Hassibi and Marzetta [11] for deriving compact integral expressions for the mutual information for isotropically random unitary transmitted source matrices S may be useful here. This is an interesting open problem.

4 Low Probability of Detection: Impact on Capacity

In this scenario the eavesdropper attempts to detect the presence of a transmitted signal against noise alone based on L observations $\mathbf{Y} = \{Y^i\}_{i=1}^L$ of his channel output. Formally, define two hypotheses $H_0 : Y^i = W^i$ and $H_1 : Y^i = S^i H^i + W^i$, $i = 1, \ldots, L$. For any strictly positive prior probabilities $P(H_0)$ and $P(H_1) = 1 - P(H_0)$ of these hypotheses, the minimum attainable probability of decision error P_e of the eavesdropper has the following large sample limiting behavior [6]:

$$\begin{aligned} \liminf_{L\to\infty} \frac{1}{L} \ln P_e &= \rho \\ \rho &= \inf_{\alpha\in[0,1]} \liminf_{L\to\infty} \frac{1}{L} \ln \int f_{H_1}^{1-\alpha}(Y_1, \ldots, Y_L) f_{H_0}^{\alpha}(Y_1, \ldots, Y_L) dY_1 \ldots dY_L \end{aligned}$$

The non-negative constant ρ is the called the Chernoff error exponent and is the error rate which determines how quickly the decision error decays exponentially to zero. This error rate is the minimum (unnormalized) α-divergence between the eavesdropper's densities f_{H_1} (the alternative density) and f_{H_0} (the null density)

which is a measure of the ease of discrimination between the two statistical distributions. This constant ρ must be negative for a patient eavesdropper to be able to correctly detect signal presence with arbitrarily low probability of error as number of time frames L increases. The objective of LPD-secure modulation is to design signaling strategies which constrain ρ to a large value (small negative value near zero if possible) and achieve highest possible information rates to the client. To this aim, we will compute the informed channel capacity of the client under such an LPD constraint for low SNR and for several eavesdropper scenarios.

4.1 SH-Informed Eavesdropper

Assume that the eavesdropper knows his channel sequence $\mathbf{H} = \{H_{TE}^i\}_{i=1}^L$ and also knows the matrix valued amplitudes $\mathbf{s} = \{s_i\}_{i=1}^L$ of the signals sequence $\mathbf{S} = \{\mathbf{S}_i\}_{i=1}^L$ transmitted over the L frames. The eavesdropper's null and alternative densities become

$$f_{H_1}(\mathbf{Y}) = f(\mathbf{Y}|\mathbf{H}, \mathbf{S} = \mathbf{s}), \qquad f_{H_0}(\mathbf{Y}) = f(\mathbf{Y}|\mathbf{H}, \mathbf{S} = 0).$$

An exact analysis of minimum probability of error P_e is possible in this Gaussian case from which it can be shown that P_e is monotone decreasing in the detectibility index that is linearly proportional to the magnitude Chernoff error exponent $|\rho(\mathbf{H}, \mathbf{s})|$. The α-divergence is simply computed

$$\ln \int f^{1-\alpha}(\mathbf{Y}|\mathbf{H}, \mathbf{S} = \mathbf{s}) f^{\alpha}(\mathbf{Y}|\mathbf{H}, \mathbf{S} = 0) d\mathbf{Y} \quad = \quad -\alpha(1-\alpha) \liminf_{L\to\infty} \sum_{i=1}^{L} \mathrm{tr}\{s_i H_i H_i^\dagger s_i^\dagger\}$$

which is minimized over $\alpha \in [0, 1]$ by the choice $\alpha = 1/2$. Thus

$$\rho = \liminf_{L\to\infty} \frac{1}{L} \sum_{i=1}^{L} \rho(H_i, s_i), \tag{8}$$

where

$$\rho(H_i, s_i) = -\frac{\eta_e^2}{4} \mathrm{tr}\{s_i H_i H_i^\dagger s_i^\dagger\}.$$

Since the eavesdropper would not normally be cooperating with the transmitter to provide feedback of his channel coefficients a reasonable LPD signaling strategy would be to try to constrain the channel-averaged Chernoff error exponent

$$\overline{\rho}(s_i) = E[\rho(H_i, s_i)|\mathbf{S} = \mathbf{s}] = -\frac{\eta_e^2}{4} \mathrm{tr}\{s_i^\dagger s_i\}.$$

For example, the transmitter could constrain the magnitude of the channel-averaged Chernoff exponent for each signal frame

$$|E[\overline{\rho}(S_i)|\mathbf{S} = \mathbf{s}]| \asymp \mathrm{tr}\{s_i^\dagger s_i\} \le P_{peak}^2, \quad i = 1, \ldots, L$$

where $x \asymp y$ denotes "x linearly proportional to y". We identify the above as an *instantaneous power constraint.*

An alternative strategy would be for the transmitter to generate i.i.d. signal matrices S_i from a source S having source distribution $dP(S)$ and satisfying the mean power constraint

$$|E[\overline{\rho}(S)]|/(TM) \asymp \mathrm{tr}\left\{\overline{S^\dagger S}\right\}/(TM) \leq P_{\mathrm{avg}} \tag{9}$$

where $\overline{S^\dagger S} = E[S^\dagger S]$. By the strong law of large numbers this is equivalent to constraining the exponent ρ (8) under the assumption that $\{H_i\}$ and $\{S_i\}_i$ are i.i.d. sequences of matrices.

Recall that under the mean power constraint $\mathrm{tr}\{\mathrm{cov}(S)\}/(TM) \leq P_{\mathrm{avg}}$ the informed channel capacity is attained by zero mean complex Gaussian S with covariance $\mathrm{cov}(S) = I_T \bigotimes \overline{S^\dagger S}$. Hence, (9) is an equivalent constraint on S and we conclude that when the eavesdropper knows both the channel and the signal, the standard mean transmit power constraint also ensures a modicum of LPD performance.

4.2 S-Informed Eavesdropper

Assume that the eavesdropper knows the signal amplitudes $\mathbf{s} = \{s_i\}_{i=1}^L$ but not the channel $H = H_{TE}$. In this case the eavesdropper's densities become

$$f_{H_1}(Y) = f(\mathbf{Y}|\mathbf{S} = \mathbf{s}), \qquad f_{H_0}(\mathbf{Y}) = f(\mathbf{Y}|\mathbf{S} = 0).$$

As both densities are multivariate Gaussian the α-divergence is again simply computed

$$\ln \int f^{1-\alpha}(\mathbf{Y}|\mathbf{S} = \mathbf{s}) f^{\alpha}(\mathbf{Y}|\mathbf{S} = 0) d\mathbf{Y} \quad = \quad \sum_{i=1}^{L} \ln \frac{|I_T + \eta_e s_i s_i^\dagger|^{1-\alpha}}{|I_T + \eta_e (1-\alpha)\ s_i s_i^\dagger|}$$

A simple asymptotic development gives

$$\ln \frac{|I_T + \eta_e s s^\dagger|^{1-\alpha}}{|I_T + \eta_e (1-\alpha)\ s s^\dagger|} \quad = \quad -\frac{\alpha(1-\alpha)\eta_e^2}{2}\ \mathrm{tr}\{s^\dagger s s^\dagger s\} + o(\eta_e^2).$$

Thus, after substituting the minimizing value $\alpha = 1/2$, the Chernoff exponent ρ has the low SNR representation

$$\rho = -\frac{\eta_e^2}{8}\ \liminf_{L\to\infty} \frac{1}{L} \sum_{i=1}^{L} \mathrm{tr}\{s_i s_i^\dagger s_i s_i^\dagger\} + o(\eta_e^2). \tag{10}$$

We conclude that for the S-informed eavesdropper and low SNR an appropriate "instantaneous LPD" constraint for the tranmsitter is

$$\mathrm{tr}\{s_i^\dagger s_i s_i^\dagger s_i\} \leq P_{peak}^4,$$

a constraint which we call the *instantaneous fourth moment* constraint.

If the transmitted signal matrices $\{s_i\}_{i=1}^L$ are i.i.d. realizations of a source S with source distribution $dP(S)$ then with probability one, from the strong law of large numbers applied to (10), constraining ρ is equivalent to constraining the *mean fourth moment* of the source

$$\frac{1}{TM}\mathrm{tr}\{\overline{S^\dagger S S^\dagger S}\} \leq P_{avg}^4, \tag{11}$$

where P_{avg}^4 is a specified constant.

4.3 Uninformed Eavesdropper

In this case the eavesdropper does not know the amplitudes $\mathbf{s} = \{s_i\}_{i=1}^L$ of the transmitted signals nor the channel H_{TE}. We will assume that $\{s_i\}$ is a realization of an i.i.d. source S for which the source distribution $dP(S)$ is known to the eavesdropper. This is a conservative assumption – in the absence of such knowledge the eavesdropper can only have worse detection error rates than predicted below.

The eavesdropper's densities are

$$\begin{aligned} f_{H_1}(\mathbf{Y}) &= f(\mathbf{Y}|\mathbf{S} \neq 0) = \int_{\mathbf{s}\neq 0} f(\mathbf{Y}|\mathbf{S}=\mathbf{s}) dP(d\mathbf{s}) \\ f_{H_0}(\mathbf{Y}) &= f(\mathbf{Y}|\mathbf{S}=0). \end{aligned}$$

The α-divergence is not closed form for the uninformed eavesdropper since it involves the difficult marginalization over $\mathbf{S}$ required for computation of f_{H_1}. However, we can apply the method of Edgeworth expansion to develop $f(Y|S \neq 0)$ about a Gaussian density to obtain the low SNR approximation (see Appendix B):

$$\begin{aligned} &\ln \int f^{1-\alpha}(Y|S\neq 0) f^{\alpha}(Y|S=0) dY \\ &\quad = \ln\left(\frac{|I_{TM} + \eta_e \mathrm{cov}(S)|^{1-\alpha}}{|I_{TM} + \eta_e (1-\alpha)\mathrm{cov}(S)|} \right) + \frac{\alpha(1-\alpha)^2 \eta_e^4}{8} \sigma_{t,u} \kappa^{t,u,v,w} \sigma_{v,w} + o(\eta_e^4) \end{aligned} \tag{12}$$

where $\mathrm{cov}(S) = ((\sigma_{t,u}))_{t,u=1}^{TM}$ is the $TM \times TM$ covariance matrix associated with the $T \times M$ source matrix S, $\kappa_{r,s,t,u} = \kappa_{r,s,t,u}(SH_{TE})$ is the $TN_E \times TN_E \times TN_E \times TN_E$ fourth order kurtosis tensor of the $T \times N_E$ matrix SH_{TE} and Einstein summation notation is used for the tensor product in the second term of (12). For a more precise definition of $\kappa_{r,s,t,u}$ see Appendix B.

As the additive eavesdropper noise W_{TE} is Gaussian the kurtosis of the eavesdropper's received signal Y satisfies $\kappa(Y) = \kappa(\sqrt{\eta_e} SH_{TE} + W_{TE}) = \eta_e^2 \kappa(SH_{TE})$. If an element of $\kappa(Y)$ is negative then the received

signal matrix has a sub-Gaussian (light-tailed) component while if an element of $\kappa(Y)$ is positive this matrix has a super-Gaussian (heavy-tailed) component. The kurtosis tensor product in (12) can be explicitly expressed as a function of the moments of the transmitted signal matrix S using the fact that the entries of H are i.i.d. complex Gaussian

$$\sigma_{t,u}\kappa^{t,u,v,w}(SH_{TE})\sigma_{v,w} \quad = \quad \eta_e^2 2N_E \sum_{k=1}^{T} \sum_{t,u,v,w=1}^{M} \mathrm{cov}(s_{kt}, s_{ku})\mathrm{cov}(s_{kt}s_{ku},\ s_{kv}s_{kw})\mathrm{cov}(s_{kv}, s_{kw}).$$

As for fixed k

$$\mathrm{cov}(s_{kt}s_{ku},\ s_{kv}s_{kw}) = E[(s_{kt}s_{ku}^* - E[s_{kt}s_{ku}^*])\ (s_{kv}s_{kw}^*] - E[s_{kv}s_{kw}^*])^*] \tag{13}$$

is a non-negative definite function in the pairs of indices (t,u) and (v,w), the kurtosis tensor product is non-negative and increases in the centralized 4th moment $\mathrm{cov}(s_{kt}s_{ku},\ s_{kv}s_{kw})$ of the source. This reflects the fact that under the assumption of a random Gaussian channel, the received signal is always super-Gaussian, i.e. it's kurtosis is greater than zero, unless the signal has zero variance.

Note that the α-divergence (12), and hence the error rate ρ, is an increasing function of the received kurtosis tensor $\kappa(Y)$ for all $\alpha \in [0,1]$. We conclude that the best countermeasure to thwart eavesdropper signal detection is for the transmitter to transmit signals leading to as high positive kurtosis of Y as possible. In particular, for fixed non-zero transmitted power, an effective LPD signaling scheme would transmit signals S having large centralized fourth moment tensor (13). This strategy may be closely related to diminishing the ability of the eavesdropper to perform blind equalization which, for the case of a scalar channel with memory, is known to be possible only when the source's fourth moment is sufficiently small to make the kurtosis negative valued [32]. The choice of a signal distribution $dP(S)$ which minimizes the α-divergence (12) or maximizes the $O(\eta_e^4)$ tensor product therein is an interesting open problem.

Considerable simplification occurs when the SNR is very low and terms of order η_e^4 can be neglected. In this case the Chernoff error exponent, i.e. the minimum of the α-divergence (12), becomes

$$\rho \quad = \quad -\frac{\eta_e^2}{8}\ \mathrm{tr}\{\mathrm{cov}^2(S)\} + o(\eta_e^2) \tag{14}$$

5 LPD-Constrained Capacity

Here we use the asymptotic error rate (14) to motivate an LPD constraint under which we derive the channel capacity C_{lpd}^{TR} for the case where both transmitter and receiver know their channel (T/R-informed) and C_{lpd}^{R} for the case that only the receiver knows the channel (R-informed).

The asymptotic LPD constraint (14) is equivalent to the *mean squared power* constraint

$$\frac{1}{TM}\sum_{i=1}^{TM}\sigma_i^2 \leq P_{\text{lpd}}, \tag{15}$$

where P_{lpd} is a prespecified maximum tolerable mean squared power and σ_i are the eigenvalues of the $TM \times TM$ matrix $\text{cov}(S)$.

5.1 T/R-Informed LPD-Capacity

In Proposition 3 in Appendix C we give the following expression for mean-square-power constrained capacity where, λ_i are the eigenvalues of $\eta_r H_{TR} H_{TR}^\dagger$ and U is a $M \times M$ unitary matrix whose columns are the (right) eigenvectors of $H_{TR} H_{TR}^\dagger$

$$\begin{aligned} C_{\text{lpd}}^{TR} &= TE\left[\ln\left|I_N + \eta_r H_{TR}^\dagger \Sigma_{\text{lpd}} H_{TR}\right|\right] \\ &= TE\left[\log\left(\frac{\sqrt{1+\mu\lambda_i^2}}{2}\right)\right]. \end{aligned} \tag{16}$$

Capacity is attained by a zero mean Gaussian source S with covariance $\text{cov}(S) = I_T \bigotimes \Sigma_{\text{lpd}}$ where $\Sigma_{\text{lpd}} = UDU^\dagger$, $D = \text{diag}(\sigma_i)$,

$$\sigma_i = \frac{\sqrt{1/\lambda_i^2 + \mu} - 1/\lambda_i}{2}, \tag{17}$$

and $\mu > 0$ is a parameter such that $M^{-1}\sum_{i=1}^{M}\sigma_i^2 = P_{\text{lpd}}$. In the sequel we will call C_{lpd}^{TR} the T/R-informed LPD-capacity and the capacity-achieving signal covariance will be called the T/R-informed LPD-optimal signal covariance.

Observe the following

1. Like the power-optimal source (6) which achieves T/R-informed power-capacity C_{pow}^{TR}, the LPD-optimal source S has i.i.d. Gaussian rows and each row has (spatial) covariance Σ_{lpd} whose eigenstructure is matched to the eigenstructure (modes) of the channel.

2. In contrast to the waterfilling strategy, distributing transmitted energy only to the highest SNR channel modes is **not** optimal for attaining the LPD-capacity.

3. The mean-squared-power and mean-power constraints can be related to each other by the Schwarz inequality

$$\sqrt{M\ \text{tr}\left\{E[S^\dagger S]E[S^\dagger S]\right\}} \geq \text{tr}\left\{E[S^\dagger S]\right\}$$

Thus the mean-squared-power constrains the mean-power of the transmitted signal. However the two constraints produce qualitatively different optimal source covariances.

4. The eavesdropper's Chernoff exponent (14) only depends on his N_E antennas through his received SNR η_e. Hence, if the eavesdropper's Chernoff exponent is to be controlled via the mean-square-power constraint (15), the transmitter's LPD-optimal signaling strategy depends on N_E only through the mean-square power constraint level P_{lpd}. In particular, if the transmitter knows that N_E has increased he will only need reduce his transmit power to ensure the same eavesdropper Chernoff exponent.

5.2 R-informed LPD Capacity

In Proposition 4 of Appendix C we establish that when the transmitter does not know the channel but the receiver does know the channel the mean-squared-power constraint (15) and the mean power constraint (4) produce the same optimal signaling strategy and result in identical forms for the channel capacity. Thus we conclude that when the eavesdropper has low received SNR his Chernoff exponent is controlled by average transmitter power and no special countermeasures are required to enhance security of the client's link.

6 Numerical Comparisons

here we compare the T/R-informed power-capacities and LPD-capacities derived in the previous section. Simulations of a Rayleigh fading channel were performed and the T/R-informed capacities under both P_{avg} and P_{lpd} constraints were computed empirically. The number N_R of transmit antennas was chosen equal to the number M of the client's receive antennas. In Fig. 3 we show the eigenvalues (diagonal entries) of the optimal signal covariance matrices which acheive each one of the capacities. These eigenvalues are indexed by the modes of the channel and are denoted as such in the figure. Both signal covariances are power normalized, i.e. they have the same trace. The LPD-optimal eigenvalue distribution is flatter and its peaks are much less prominent than the standard power-optimal eigenvalue distribution. This reflects the intuitive fact that an eavesdropper can less easily detect the presence of a flat signal eigenvalue profile and hence the LPD-optimal signaling strategy better hides the signal information than the power-optimal signaling strategy.

In Figs. 4 and 5 are plotted the T/R-informed standard power-capacity and LPD-capacity as a function of SNR $\eta = \eta_r$ for various numbers of antenna elements ($N_R = M$).

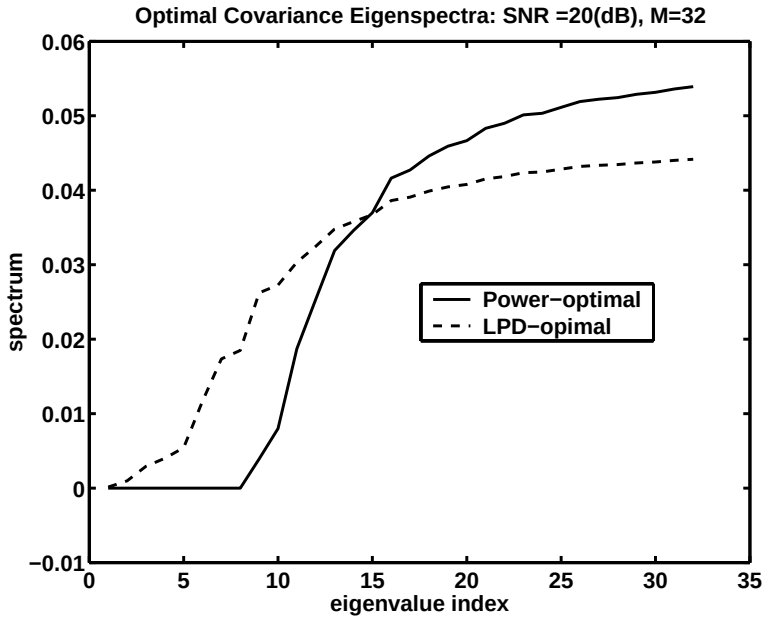

Figure 3: *Spectra of optimal source covariance matrices under T/R-informed LPD (mean-squared-power) and mean-power constraints:* $SNR = 20dB, M = N_R = 32$.

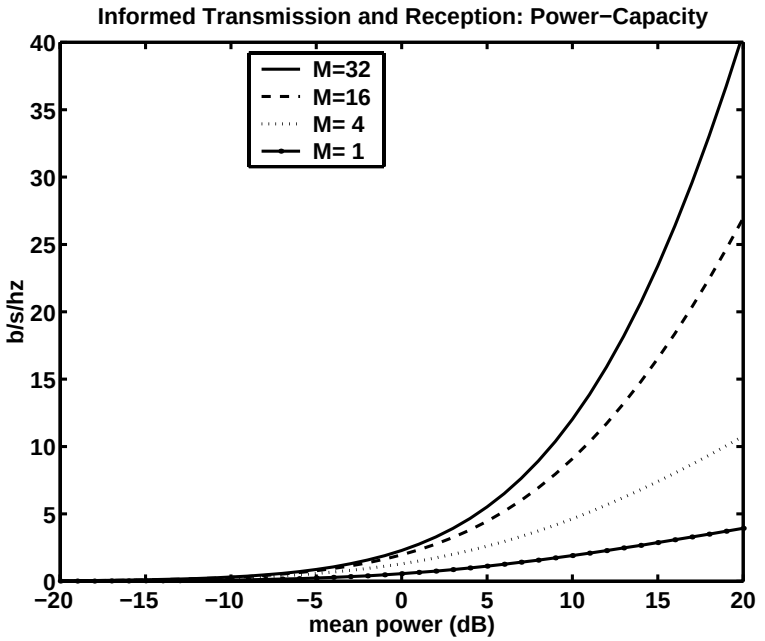

Figure 4: *T/R informed power-capacity* C_{pow}^{TR} $(N_R = M)$

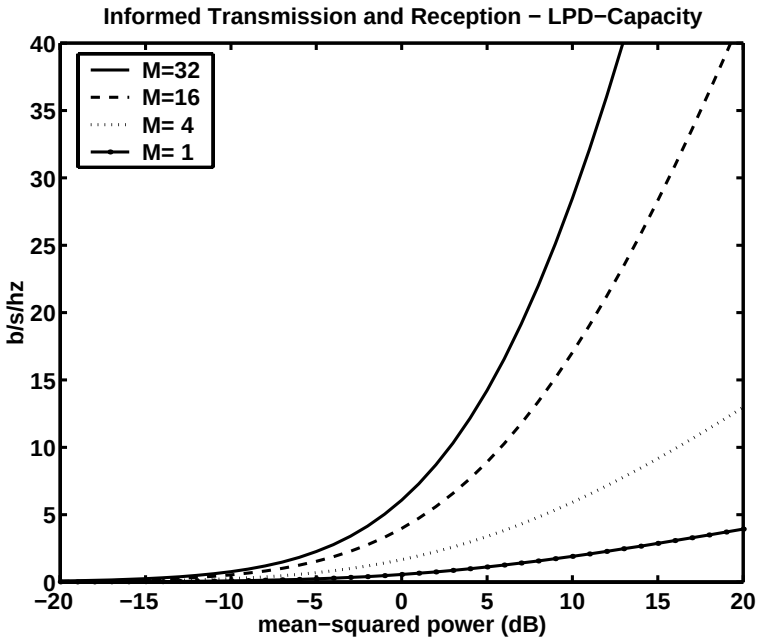

Figure 5: *T/R informed LPD-capacity* $C_{\rm lpd}^{TR}$ *(*$N_R = M$*)*

Next we investigated tradeoffs between the T/R-informed LPD-optimal signaling strategy and the standard power-optimal waterfilling signaling strategy. Define

$$\mathcal{I}_c(\Sigma) = TE\left[\ln\left|I_N + \eta_r H_{TR}^{\dagger}\Sigma H_{TR}\right|\right]$$

the average information rate attained by a zero mean Gaussian source S with covariance matrix $\mathrm{cov}(S) = I_T \bigotimes \Sigma$ satisfying the constraint denoted by c. The notation means that if $c = P_{\rm avg}$ then Σ satisfies $\mathrm{tr}\{\Sigma\}/(TM) \le P_{\rm avg}$ and if $c = P_{\rm lpd}$ then it satisfies $\mathrm{tr}\{\Sigma^2\}(TM) \le P_{\rm lpd}$. Thus the standard power-capacity is $C_{\rm pow}^{TR} = \mathcal{I}_{P_{\rm avg}}(\Sigma_{\rm pow})$ where the power-optimal signal covariance $\Sigma_{\rm pow}$ is specified by (6) and the LPD-capacity is $C_{\rm lpd}^{TR} = \mathcal{I}_{P_{\rm lpd}}(\Sigma_{\rm lpd})$ where LPD-optimal signal covariance $\Sigma_{\rm lpd}$ is specified by (17).

The loss in power-capacity due to using the LPD-optimal signal covariance structure $\Sigma_{\rm lpd}$ is defined as

$$\mathcal{I}_{P_{\rm avg}}(\Sigma_{\rm lpd})/\mathcal{I}_{P_{\rm avg}}(\Sigma_{\rm pow}) \tag{18}$$

while the loss in LPD-capacity due to using the power-optimal signal covariance structure $\Sigma_{\rm pow}$ is

$$\mathcal{I}_{P_{\rm lpd}}(\Sigma_{\rm pow})/\mathcal{I}_{P_{\rm lpd}}(\Sigma_{\rm lpd}) \tag{19}$$

In (18) both the LPD-optimal and the power-optimal covariances are forced to satisfy the same mean-power constraint while in (19) they both satisfy the same mean-squared-power constraint. Figures 6 and 7 plot the capacity losses as a function of mean-power and mean-squared-power, respectively. Notice that the

loss increases as more antennas $M = N_R$ are deployed by transmitter and client. This is because a higher proportion of the signal covariance eigenspectrum is flattened out by the LPD-optimal signaling strategy as compared to the power-optimal strategy. Also note that as the client's SNR η_r increases the relative capacity loss becomes negligible while as η_r decreases to -20 dB the losses flatten out. This is because at very low η_r the power-optimal waterfilling strategy requires the transmitter to use only a single transmit antenna while the LPD-optimal signal applies energy to all antennas no matter how low η_r gets. Finally for a single transmitter antenna element ($M = 1$) there is no loss in capacity since in this case the average power and the mean-squared-power constraints are equivalent up to a scale factor.

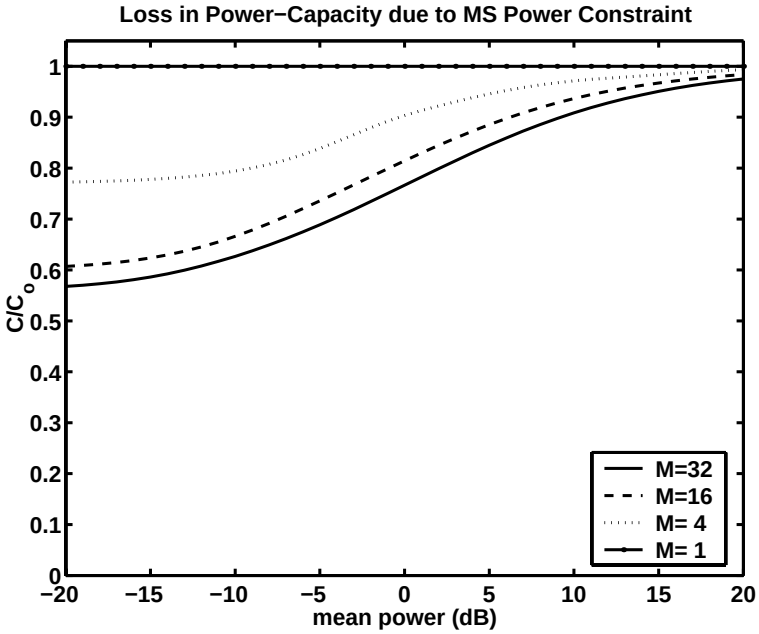

Figure 6: *Loss in power-capacity due to mean-squared power (LPD) constraint ($N_R = M$)*

Finally, we investigated the sensitivity of the T/R-informed power-capacity and LPD-capacity due to errors in the transmitter's channel estimate. The receiver is assumed to have decode symbols with zero channel estimation error. This asymmetric channel error scenario is an idealization of the situation where channel estimation errors occur during training which are then fedback to the transmitter. While we offer no proof, we believe that the effect on capacity of using erroneous channel information at the transmitter is greater than using equivalent error estimates at the receiver and therefore these results should approximate actual information rate reductions due to training. This would have to verified by doing more extensive simulations to determine the mutual information loss due to training errors at both transmitter and receiver. In our simulation the total number of samples in a coherent fade was $T = 1024$ and 128 of these samples

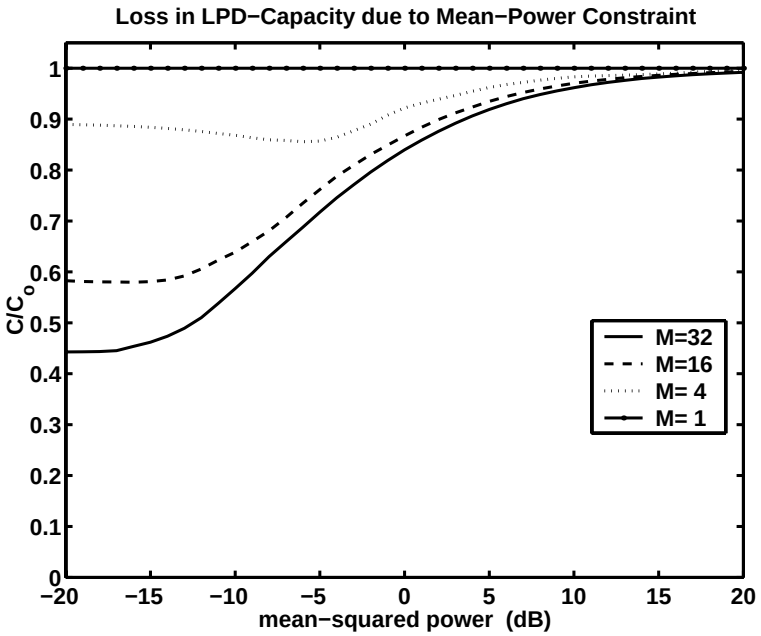

Figure 7: *Loss in LPD-capacity due to mean-power constraint ($N_R = M$)*

were used for estimating the channel at the receiver. 500 realizations of different Gaussian channels were generated for various numbers of antennas and SNRs. Over each frame the channel was estimated at the receiver via the exact least squares estimator based on 128 N_R-element snapshots generated by transmitted zero mean i.i.d. Gaussian training symbols. These channel estimates were then substituted into the power-optimal and LPD-optimal covariances $\Sigma_{\rm pow}$ and $\Sigma_{\rm lpd}$ and substituted into the capacity equations (5) and (16), respectively. Figs. 8 and 9 shows the resultant degradation in these two capacities. Observe that for the example simulated here, for moderate to large SNR the relative loss due to transmitter-channel mismatch is significantly less than the loss due to not accounting for the eavesdropper LPD constraint (compare Fig. 9 to Fig. 7).

7 Conclusions

This paper has presented a study of capacity under link security constraints corresponding to low probability of intercept (LPI) and low probability of detect (LPD). We have established that optimal signaling for LPD- and LPI- constrained *secure* channels is qualitatively different from *open* channels. We have also shown that constraining moment quantites, such as trace of 4th moment matrix, are relevant for eluding detection by eavesdroppers who have only limited knowledge about the channel and transmitter modulation. A smart eavesdropper with info on data or training sequences can be handled similarly by constraining the fourth

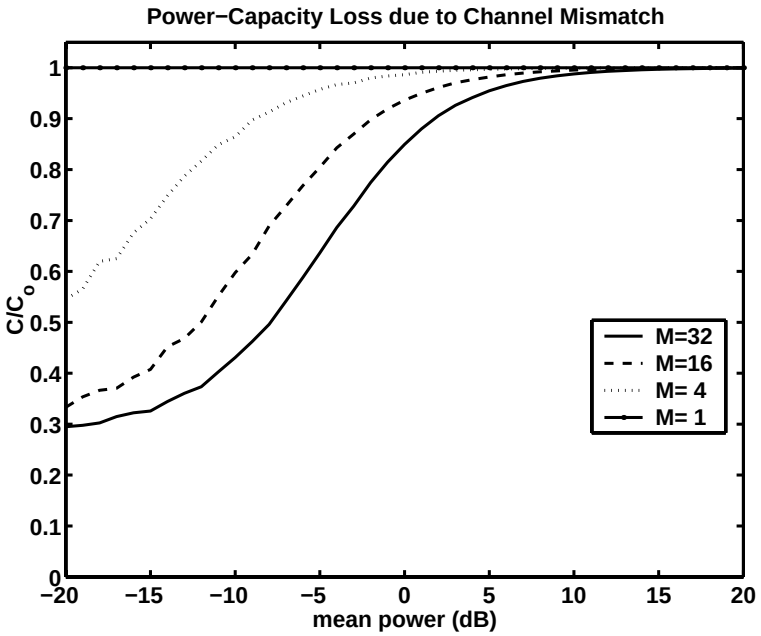

Figure 8: *T/R informed power-capacity loss due to transmitter-channel mismatch ($N_R = M$)*

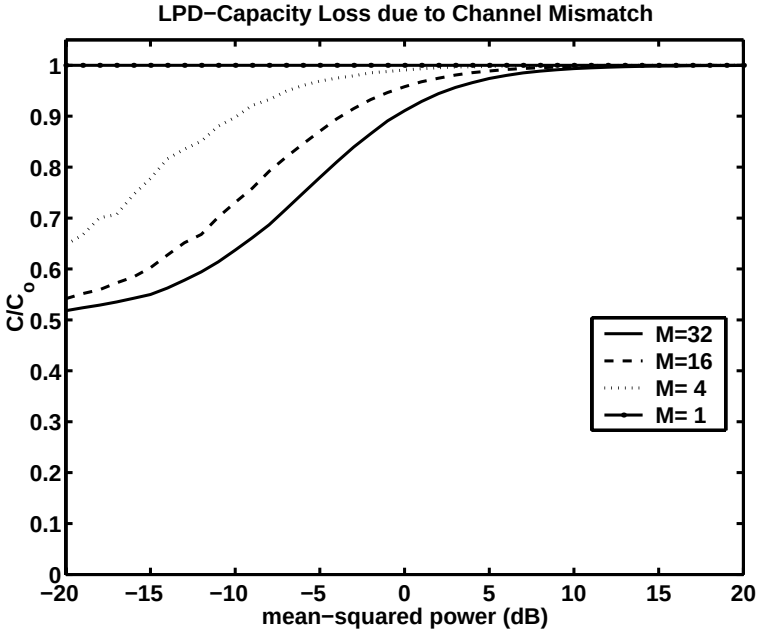

Figure 9: *T/R informed LPD-capacity loss due to transmitter-channel mismatch ($N_R = M$)*

moment of the transmitted signal matrices. The analysis in this paper holds only for doubly informed links for which the receiver and transmitter know their channel exactly. Extensions of these results to the case of network-wide QoS metrics such as min- and sum- capacity are of interest. This paper has treated the case of Gaussian noise and known receiver noise covariance matrices. Generalizations to the case of non-Gaussian multi-user interference (MUI) would be worthwhile for answering questions such as: to what extent are LPD and MUI resistance compatible goals in wireless networks. Finally, another interesting avenue for exploring the information hiding capabilities of space-time channels would be the information theoretic framework of Moulin [24, 23].

Appendix A

Proposition 1 *Assume that H is an $M \times N$ matrix of i.i.d. zero mean and unit variance circularly symmetric Gaussian elements. If the density dP^* defined below exists, the T/R-informed perfect-secrecy capacity is $E[C^{TR}_{\text{perfsec}}(H)]$ where*

$$C^{TR}_{\text{perfsec}}(H) = -\int dX \mathcal{N}_{TN}(SH - X) \ln \int_{\mathcal{S}} dP^*(S') \mathcal{N}_{TN}(S'H - X) - TN \ln \pi e \tag{20}$$

where

- *$\mathcal{N}_{TN}(X)$ is the probability density of a standard zero mean and identity covariance $T \times N$ complex normal matrix.*
- *dP^*_S is a probability density over the perfect secrecy constraint set $\mathcal{S} = \{S : SS^\dagger = A\}$, A a nonrandom matrix, which makes the RHS of (21) functionally independent of S over $\mathcal{S}$*

If $N \geq M$ then, with $H = U[\Lambda\ O]V$ the SVD of H

$$C^{TR}_{\text{perfsec}}(H) = -\int dX_1 \mathcal{N}_{TM}(SU\Lambda - X_1) \ln \int_{\mathcal{S}} dP^*_S(S') \mathcal{N}_{TM}(S'U\Lambda - X_1) - TM \ln \pi e \tag{21}$$

where the $T \times M$ matrix X_1 is defined by the following column partition of the matrix XV: $[X_1, X_2] = XV$. Furthermore, if $\Lambda = \sigma I_M$ for $\sigma > 0$ then $dP^*(S) = 1/\text{SO}(\mathbb{C}^{TM})$ is the uniform density over $\mathcal{S}$.

Proof:

For fixed H the T/R-informed perfect secrecy capacity is defined as $C^{TR}_{\text{perfsec}}(H) = \sup_{dP(S)} \mathcal{I}(X, S|H)$ where the maximization is over distributions supported on $\mathcal{S}$. As $\mathcal{I}(X, S|H) = \mathcal{H}(X|H) - \mathcal{H}(X|S, H) = \mathcal{H}(X|H) - TN \log \pi e$ we focus on the entropy function $\mathcal{H}(X|H)$

$$\begin{aligned} \mathcal{H}(X|H) &= -\int_X dX f(X|H) \log f(X|H) \\ &= -\int_{\mathcal{X}} dX \int_{\mathcal{S}} f(X|S, H) dP(S) \log \int_{\mathcal{S}} f(X|S, H) dP(S) \end{aligned}$$

where $\mathcal{X} = \mathbb{C}^{TN}$. As $\mathcal{H}(X|H)$ is concave as a function of $dP(S)$ a standard calculus of variations argument can be used to derive a sufficient condition for the maximum of $\mathcal{H}(X|H)$. Define the Lagrangian

$$L(dP(S)) = \mathcal{H}(X|H) - \gamma \int_{\mathcal{S}} dP(S)$$

where γ is an undetermined positive multiplier which enforces the constraint that $\int_{\mathcal{S}} dP(S) = 1$. The stationary point condition on the maximizing $dP(S) = dP^*(S)$ follows from considering

$$L(dP^*(S) + d\Delta(S)) - L(dP^*(S)) \quad = \quad \int_{\mathcal{S}} d\Delta(S) \left[\int_{\mathcal{X}} dX f(X|S,H) \ln \int_{\mathcal{S}} f(X|S,H) dP^*(S) - \gamma \right].$$

For the above to be zero for all $d\Delta(S)$ such that $dP^*(S) + d\Delta(S)$ remains a valid distribution over $\mathcal{S}$ we require

$$\int_{\mathcal{X}} dX f(X|S,H) \ln \int_{\mathcal{S}} f(X|S,H) dP^*(S) = \gamma, \quad \forall\, S \ \in \mathcal{S}$$

which, upon substituting $f(X|S,H) = \mathcal{N}_{TN}(SH - X)$, gives the expression (20) of the Proposition.

Specializing to the case that $N \geq M$, and using the given SVD of H and definitions of X_1 and X_2, observe that

$$\begin{aligned} \mathcal{N}_{TN}(SH - X) &= \frac{1}{\pi^{TN}} \exp\left(-\mathrm{tr}\{[SH - X][SH - X]^\dagger\}\right) \\ &= \frac{1}{\pi^{TN}} \exp\left(-\mathrm{tr}\{[SU\Lambda - X_1][SU\Lambda - X_1]^\dagger\} - \mathrm{tr}\{X_2 X_2^\dagger\}\right) \\ &= \mathcal{N}_{TM}(SU\Lambda - X_1)\, \mathcal{N}_{T(N-M)}(X_2) \end{aligned}$$

Plugging this back into the expression for $\mathcal{I}(X, S|H)$ and using simple algebra this establishes (21).

Finally, assuming $\Lambda = \sigma I_M$, from expression (21) we must show that

$$g(S) \stackrel{\mathrm{def}}{=} \int dX_1 \mathcal{N}_{TM}(SU - X_1/\sigma) \ \ln \int_{\mathcal{S}} dS' \mathcal{N}_{TM}(S'U - X_1/\sigma)$$

is equal to a constant over $\mathcal{S}$ independent of S. Note that $\mathcal{N}_{TM}(SU - X_1) = \mathcal{N}_{TM}(S - X_1 U^\dagger)$ and observe that, by change of variable in the integrals above, for any $M \times M$ unitary matrix Φ: $g(S\Phi) = g(S)$. Hence, by Vinograd's theorem, g only depends on S through $SS^\dagger$. Since $SS^\dagger = A$ over $\mathcal{S}$, $g(S)$ must in fact be constant over $\mathcal{S}$ and the optimality of the uniform distribution dP^* is established. $\square$

Appendix B

As in Section 2 let $Y = \sqrt{\eta} SH + W$, be an $T \times N$ matrix of complex amplitudes measured at N antennas over T time samples at the output of an i.i.d. zero mean complex Gaussian $M \times N$ channel H with additive complex Gaussian noise W and input signal (source) with $T \times M$ complex amplitudes S. While cooperation between client and transmitter may lead to dependency between S and H, we assume there is no such cooperation between eavesdropper and transmitter. Thus, as the results below are applied to the eavesdropper, we assume here that S is independent of W and H.

Let $Y = Y_R + jY_I$ be a complex valued $T \times N$ matrix where Y_R and Y_I are real matrices. Define the real valued $2TN$-element vector $\vec{Y}$ as the concatenation of the $2N$ columns of the matrix $[Y_R, Y_I]$. Following the notation of [22], for zero mean Y, define the tensors $\kappa_{r,s}$, $\kappa_{r,s,t}$ and $\kappa_{r,s,t,u}$ as the variance, skewness, and kurtosis of Y ($\vec{Y}$):

$$\begin{aligned}
\kappa_{r,s}(Y) &= E[\vec{Y}_r \vec{Y}_s] \\
\kappa_{r,s,t}(Y) &= E[\vec{Y}_r \vec{Y}_s \vec{Y}_r] \\
\kappa_{r,s,t,u}(Y) &= E[\vec{Y}_r \vec{Y}_s \vec{Y}_r \vec{Y}_s] - E[\vec{Y}_r \vec{Y}_s] E[\vec{Y}_r \vec{Y}_s] \quad [3].
\end{aligned}$$

Note that as W is Gaussian $\kappa_{r,s,t,u}(Y) = \kappa_{r,s,t,u}(SH_{TE}) = O(\eta^2)$. Further, define λ_{rs} the rs element of the inverse of the covariance matrix $\kappa_{t,s}(Y)$ and ϕ_{ts} the positive definite square root factor of λ_{ts}, i.e., using the Einstein summation convention $\lambda^{rs}\kappa_{s,t} = \delta_{rt}$, the kronecker delta function, and $\phi^{tu}\phi_{us} = \lambda_{ts}$.

Proposition 2 *Assume that S is independent of W and H and that W and H are mutually independent zero mean complex Gaussian matrices. The (normalized) α-divergence between $f(Y|S \neq 0)$ and $f(Y|S = 0)$ has the asymptotic expression*

$$\ln \int f^{1-\alpha}(Y|S \neq 0) f^{\alpha}(Y|S = 0) dY \tag{22}$$

$$= \ln \frac{\left| I_T + \eta \overline{SS^\dagger} \right|^{1-\alpha}}{\left| I_T + \eta(1-\alpha)\overline{SS^\dagger} \right|} + \frac{\alpha(1-\alpha)^2\eta^2}{8} \tilde{\sigma}_{t,u} \tilde{\kappa}_{t,u,v,w} \tilde{\sigma}_{v,w} + o(\eta^2 \tilde{\kappa}_{t,u,v,w}) \tag{23}$$

Where $\overline{SS^\dagger} = E[SS^\dagger]$, $\tilde{\kappa}_{t,u,v,w}$ is the kurtosis of the whitened and variance normalized measurement $\phi^{st}\vec{Y}_t$ and $\tilde{\sigma}_{t,u}$ is the variance of the prewhitened signal $\phi^{st}\vec{S}_t$.

Proof

Using the expression in [22] for the Edgeworth expansion of a zero mean multivariate density about a Gaussian multivariate density with zero mean and covariance $\kappa_{r,s} = \kappa_{r,s}(Y)$ we have the representation

$$f(Y|S \neq 0) = \mathcal{N}_{2TN}(Y; 0, \kappa_{r,s}) \left(1 + \frac{1}{6}\kappa^{r,s,t} h_{rst}(Y, \lambda) + \frac{1}{72}\left[3\kappa^{r,s,t,u} h_{rstu}(Y, \lambda) + \kappa^{r,s,t}\kappa^{u,v,w} h_{rstuvw}(Y, \lambda) \right] \right) + o(\kappa^{r,s,t,u})$$

where $\mathcal{N}_{2TN}(x; 0, \kappa_{r,s})$ is the $2TN$-variate Gaussian density with zero mean and covariance $\kappa_{r,s}$, and h_{rst}, h_{rstu}, etc, are Hermite tensors given in [22]. As $Y = \sqrt{\eta}SH + W$ and H, W are independent zero mean Gaussian random matrices, and S is independent of H and W: $\kappa^{u,v,w} = \kappa^{u,v,w}(Y) = 0$ and the representation reduces to

$$f(Y|S \neq 0) = \mathcal{N}(Y; 0, \lambda^{-1}) \left(1 + \frac{1}{24}\kappa^{r,s,t,u} h_{rstu}(Y, \lambda) \right) + o(\kappa^{r,s,t,u}(Y))$$

Using $f(Y|S \neq 0) = \mathcal{N}_{2TN}(Y; 0, \delta_{rs})$, after some algebra the substitution of the Edgeworth expansion into the α-divergence expression gives

$$\ln \int f^{1-\alpha}(Y|S \neq 0) f^{\alpha}(Y|S = 0) dY \tag{24}$$
$$\ln \int \mathcal{N}(Y; 0, \lambda^{rt}_{1-\alpha}\kappa_{t,s}) \left(1 + \frac{1}{24}\kappa^{r,s,t,u} h_{rstu}(Y, \lambda)\right)^{\alpha} dY + R(\alpha)$$

where $\lambda^{rs}_{1-\alpha}$ is the matrix inverse of $\kappa^{r,s}_{1-\alpha} = \delta_{r,s} + (1-\alpha)\sigma_{r,s}$ and

$$R(\alpha) = \ln \frac{\left|I_T + \eta \overline{SS^{\dagger}}\right|^{1-\alpha}}{\left|I_T + \eta(1-\alpha)\overline{SS^{\dagger}}\right|}$$

Apply the small argument formula $(1+x)^{\alpha} \approx 1 + \alpha x$ to perform the integration (24)

$$\ln \int \mathcal{N}(Y; 0, \lambda^{rt}_{1-\alpha}\kappa_{ts}) \left(1 + \frac{1}{24}\kappa^{r,s,t,u} h_{rstu}(Y, \lambda)\right)^{\alpha} dY$$
$$= \frac{\alpha \kappa^{r,s,t,u}}{24} \int \mathcal{N}(Y; 0, \lambda^{rt}_{1-\alpha}\kappa_{ts}) h_{rstu}(Y, \lambda) + o(\eta_e^2 \tilde{\kappa}_{r,s,t,u})$$

Next we use the Hermite tensor expression [22]

$$h_{rstu}(Y, \lambda) = \vec{Y}^{'}_r \vec{Y}^{'}_s \vec{Y}^{'}_t \vec{Y}^{'}_u - \vec{Y}^{'}_r \vec{Y}^{'}_s \lambda_{tu}[6] + \lambda_{rs}\lambda_{tu}[3],$$

where $\vec{Y}^{'}_r = \lambda^{rs}\vec{Y}_s$:

$$\begin{aligned}\int \mathcal{N}(Y; 0, \lambda^{vw}_{1-\alpha}\kappa_{wx}) h_{rstu}(Y, \lambda) \quad &= \lambda_{rv}\lambda_{sw}\lambda_{tx}\lambda_{uy}\left(\kappa^{v,z_1}_{1-\alpha}\kappa^{w,z_2}_{1-\alpha}\kappa_{z_1,z_2}\kappa^{x,y}[3] - \kappa^{v,z_1}_{1-\alpha}\kappa_{z_1,w}\kappa^{xy}[6] + \kappa_{vw}\kappa_{xy}[3]\right) \\ &= (1-\alpha)^2 \lambda_{rv}\lambda_{sw}\lambda_{tx}\lambda_{uy}\sigma_{vz_1}\sigma_{xz_2}\kappa^{z_1,w}\kappa^{z_2,y}[3].\end{aligned}$$

Substituting this back into (24) and noting that, as tensors $\phi_{rs}, \kappa_{rs}, \lambda_{rs}$ have the same eigenvectors they commute,

$$\begin{aligned}\kappa^{r,s,t,u}\lambda_{rv}\lambda_{sw}\lambda_{tx}\lambda_{uy}\sigma_{vz_1}\sigma_{xz_2}\kappa^{z_1,w}\kappa^{z_2,y}[3] &= \kappa^{r,s,t,u}\phi_{rv}\phi_{sw}\phi_{tx}\phi_{uy}\ \sigma^{rz_1}\phi_{z_1 s}\sigma^{tz_2}\phi^{z_2 u} \\ &= \tilde{\kappa}^{r,s,t,u}\tilde{\sigma}^{rs}\tilde{\sigma}^{tu}[3]\end{aligned}$$

This establishes (23). □

Noting that $\tilde{\kappa}^{r,s,t,u}(Y) = \eta^2 \kappa^{r,s,t,u}(SH) + o(\eta^2)$ and $\tilde{\sigma}_{vw} = \sigma_{vw} + o(\eta)$ so that

Corollary 1

$$\ln \int f^{1-\alpha}(Y|S \neq 0) f^{\alpha}(Y|S = 0) dY$$
$$= \ln \frac{\left|I_T + \eta \overline{SS^{\dagger}}\right|^{1-\alpha}}{\left|I_T + \eta(1-\alpha)\overline{SS^{\dagger}}\right|} + \frac{\alpha(1-\alpha)^2\eta^4}{8}\ \sigma_{tu}\kappa^{t,u,v,w}\sigma_{vw} + o(\eta^4) \tag{25}$$

where $\kappa^{t,u,v,w} = \kappa^{t,u,v,w}(SH)$ is the kurtosis tensor of the $T \times N$ matrix SH.

Appendix C

Define $\mathcal{I}(S,X|H) = E[\ln f(X|S,H)/f(X|H)|H]$ the mutual information for a T/R informed link over a $M \times N$ quasi-static Rayleigh channel which is constant over the coherent fade interval of T time samples. The LPD-capacity $C_{\rm lpd}^{TR}$ of this link is defined as

$$E[\max_{f(S)} \mathcal{I}(S,X|H)]$$

where the expectation is over H and the maximization is over source distributions $dP(S)$ which satisfy the mean-squared-power constraint

$$\frac{1}{TM}\mathrm{tr}\{\mathrm{cov}^2(S)\} \leq P_{\rm lpd} \tag{26}$$

where $\mathrm{cov}(S) = E[\vec{S}^\dagger \vec{S}] - E[\vec{S}^\dagger]E[\vec{S}]$ is the source's $TM \times TM$ covariance matrix and $\vec{S}$ is a TM-element row vector constructed by concatenating the T rows of S.

Let $\eta HH^\dagger$ have eigendecomposition $\eta HH^\dagger = U\Lambda U^\dagger$ where $\Lambda = \mathrm{diag}(\lambda_i)$.

Proposition 3 *Let H be a $M \times N$ channel matrix with zero mean and unit variance i.i.d. complex circularly symmetric Gaussian entries. For the case that both transmitter and receiver know H the channel capacity under the mean-squared power constraint (26) is*

$$\begin{aligned} C_{\rm lpd}^{TR} &= TE\left[\ln\left|I_N + \eta H^\dagger \Sigma_{\rm lpd} H\right|\right] \\ &= TE\left[\log\left(\frac{\sqrt{1+\mu\lambda_i^2}}{2}\right)\right] \end{aligned} \tag{27}$$

which is attained by a zero mean Gaussian source S with covariance $\mathrm{cov}(S) = I_T \bigotimes \Sigma_{\rm lpd}$ where $\Sigma_{\rm lpd} = UDU^\dagger$, $D = \mathrm{diag}(\sigma_i)$,

$$\sigma_i = \frac{\sqrt{1/\lambda_i^2 + \mu} - 1/\lambda_i}{2},$$

and $\mu > 0$ is a parameter such that $\sum_{i=1}^M \sigma_i^2/M = P_{\rm lpd}$.

Proof

The argument is similar to that used in [33] in proving optimality of the water-filling solution for the case of informed transmitter and receiver under a mean-power constraint. The matrix observation model over a single frame $X = \sqrt{\eta}SH + W$ has the equivalent vectorized form

$$\vec{X} = \sqrt{\eta}\vec{S}\mathbf{H} + \mathcal{W} \tag{28}$$

where $\vec{X} = [X_{1*}, \ldots, X_{N*}]$ is a TN-element row vector from concatenating rows $\{X_{i*}\}_{i=1}^{T}$ of X and similarly for $\mathcal{W}$, $\vec{S} = [S_{1*}, \ldots, S_{M*}]$ is a similarly defined TM-element row vector, and $\mathbf{H} = I_T \bigotimes H$ is a block diagonal $MT \times NT$ matrix with identical diagonal blocks H. Invoking the maximum entropy property of the Gaussian distribution for $\vec{S}$ having fixed covariance $\mathrm{cov}(\vec{S}) = Q$, we have the following inequality

$$\mathcal{I}(S, X|H) \quad = \quad \mathcal{H}(X|H) - \mathcal{H}(X|S, H) \leq \log\left|I_{TN} + \eta \mathbf{H}^\dagger Q \mathbf{H}\right| \tag{29}$$

with equality when S is a zero mean complex Gaussian vector with $TM \times TM$ covariance matrix $\mathrm{cov}(S)$. It remains to maximize the right hand side of this inequality over non-negative definite symmetric matrices Q subject to $\mathrm{tr}\{Q^2\}/(MT) \leq P_{\mathrm{lpd}}$.

With the eigendecomposition $\eta HH^\dagger = U\Lambda U^\dagger$ the eigendecomposition of $\eta \mathbf{H}\mathbf{H}^\dagger$ is simply

$$\eta \mathbf{H}\mathbf{H}^\dagger = (I_T \bigotimes \mathcal{U})(I_T \bigotimes \Lambda)(I_T \bigotimes \mathcal{U})^\dagger$$

Let $\mathbf{\Sigma} = (I_T \bigotimes \mathcal{U})^\dagger Q (I_T \bigotimes \mathcal{U})$ have diagonal elements $\{\sigma_i\}_{i=1}^{TM}$. Then, by Hadamard's inequality:

$$\begin{aligned} \log\left|I_{TN} + \eta \mathbf{H}^\dagger \mathrm{cov}(S) \mathbf{H}\right| &= \log\left|I_{TN} + \Lambda \mathbf{\Sigma}\right| \\ &\leq \log \prod_{i=1}^{TM} (1 + \sigma_i \lambda_{i\%M}) \\ &= \sum_{j=1}^{T} \sum_{i=1}^{M} \log(1 + \sigma_{(j-1)T+i} \lambda_i) \end{aligned} \tag{30}$$

with equality when $\mathbf{\Sigma} = \mathrm{diag}(\sigma_i)$. Note that $\mathrm{tr}\{\mathbf{\Sigma}^2\} = \mathrm{tr}\{Q^2\} = \sum_{i=1}^{MT} \sigma_i^2$. The maximizer of the right hand side of the above equation subject to the inequality constraint $\sum_{i=1}^{TM} \sigma_i^2/(MT) \leq P_{\mathrm{lpd}}$ achieves the constraint with equality as the expression (30) is increasing in σ_i. The Lagrangian for this constrained optimization problem is

$$L(\sigma) = \sum_{j=1}^{T} \sum_{i=1}^{M} \log(1 + \sigma_{(j-1)T+i} \lambda_i) - \gamma \sum_{j=1}^{T} \sum_{i=1}^{M} \sigma_{(j-1)T+i}^2$$

where $\gamma > 0$ is an undetermined multiplier. This concave function has a unique unconstrained maximum which occurs when $0 = \partial L/\partial \sigma_{(j-1)T+i} = 1/(1 + \lambda_i \sigma_{(j-1)T+i}) - 2\gamma \sigma_{(j-1)T+i}$, or equivalently $\lambda_i \sigma_{(j-1)T+i}^2 + \sigma_{(j-1)T+i} - 1/(2\gamma) = 0$. There is one positive root $\sigma_{(j-1)T+i} = (-1 + \sqrt{1 + \lambda_i \mu})/(2\lambda_i)$ where $\mu = 2/\gamma$. Thus $\mathbf{\Sigma} = I_T \bigotimes \Sigma_{\mathrm{lpd}}$, the optimum source covariance is $Q = I_T \bigotimes (U^\dagger \Sigma_{\mathrm{lpd}} U)$, and plugging this into (30), the capacity is (27) as claimed. □

Proposition 4 *Let H be a $M \times N$ channel matrix with zero mean and unit variance i.i.d. complex circularly symmetric Gaussian entries. For the case that the only the receiver knows H the channel capacity under*

the mean-squared power constraint (26) is identical to the standard mean power constrained capacity for this case

$$C^R_{\text{lpd}} \quad = \quad TE\left[\ln\left|I_N + \sqrt{P_{\text{lpd}}}\eta H^\dagger H\right|\right] \tag{31}$$

which is attained by a $T \times M$ source matrix S whose elements are zero mean i.i.d. circularly symmetric Gaussian random variables with variances $\sqrt{P_{\text{lpd}}}$.

Proof

The proof parallels the proof of the mean-power constrained capacity in [33]. The capacity for the case that only the receiver knows the channel H is defined as $\sup_{dP(S)} E[\mathcal{I}(S,X|H)]$. Using the vectorized signal representation (28) and (29) obtained in the proof of Proposition 3 we have:

$$E[\mathcal{I}(S,X|H)] \quad \leq \quad E\left[\log\left|I_{TN} + \eta\mathbf{H}^\dagger \text{cov}(S)\mathbf{H}\right|\right] \tag{32}$$

where equality is achieved when S is zero mean circularly symmetric complex Gaussian with $TM \times TM$ covariance matrix $\text{cov}(S)$. As in [33], for any $TM \times TM$ matrix Q the function $\Psi(Q) = E\left[\log\left|I_{TN} + \eta\mathbf{H}^\dagger Q\mathbf{H}\right|\right]$ is concave and for any $MT \times MT$ unitary matrix U, $\psi(U^\dagger QU) = \psi(Q)$. Thus, specializing U to the eigenvector matrix in the eigendecomposition $\text{cov}(S) = UDU^\dagger$, we have $\psi(\text{cov}(S)) = \psi(D)$ so that, as $\text{tr}\{\text{cov}^2(S)\} = \text{tr}\{D^2\}$, without loss of generality we can assume that the capacity achieving covariance $\text{cov}(S)$ is a non-negative diagonal matrix D. Next specializing U to Π, a $TM \times TM$ permutation matrix, $\Psi(\Pi^\dagger D\Pi) = \Psi(D)$. Thus by Jensen's inequality, summing over all $(TM)!$ permutation matrices

$$\Psi(D) = \frac{1}{(TM)!}\sum_{\Pi}\Psi(\Pi^\dagger D\Pi) \leq \Psi(\tilde{D})$$

where $\tilde{D} = \frac{1}{(TM)!}\sum_\Pi \Pi^\dagger D\Pi$ which is a scaled $TM \times TM$ identity matrix. It follows from the inequality below that the constraints $\text{tr}\{D^2\}/(TM) \leq P_{\text{lpd}}$ and $\text{tr}\{\tilde{D}^2\}/(TM) \leq P_{\text{lpd}}$ are equivalent so that $\tilde{D} = \sqrt{P_{\text{lpd}}}I_{TM}$ is the optimal source covariance.

$$\begin{aligned}
\text{tr}\{\tilde{D}^2\} \quad &= \quad \sum_{k=1}^{TM}\left(\frac{1}{(TM)!}\right)^2 \sum_\Pi\sum_{\bar{\Pi}}\lambda_{\pi(k)}\lambda_{\bar{\pi}(k)} \\
&= \quad \sum_{k=1}^{TM}\left(\frac{1}{(TM)!}\sum_\Pi \lambda_{\pi(k)}\right)^2 \\
&= \quad \sum_{k=1}^{TM}\left(\frac{1}{TM}\sum_{i=1}^{TM}\lambda_i\right)^2 \\
&\leq \quad \sum_{k=1}^{TM}\lambda_k^2 = \text{tr}\{D^2\}.
\end{aligned}$$

Where the last line follows from Jensen's inequality. This establishes the proposition □

References

[1] E. Biglieri, J. Proakis, and S. S. (Shitz), "Fading channels: information-theoretic and communications aspects," *IEEE Trans. on Inform. Theory*, vol. 44, pp. 2619–2692, Oct., 1998.

[2] D. W. Bliss, K. W. Forsythe, A. O. Hero, and A. L. Swindelhurst, "MIMO environmental capacity sensitivity," in *Proc. of 34th Asilomar Conf. on Signals, Systems and Computers*, Monterey, CA, 2000. `http://www.eecs.umich.edu/~hero/comm.html`.

[3] C. Busch, W. Funk, and S. Wolthusen, "Digital watermarking: from concepts to real-time video applications," *IEEE Computer Graphics and Applications*, vol. 19, no. 1, pp. 25–35, Jan.-Feb. 1999.

[4] I. Csiszár and J. Korner, "Broadcast channels with confidential messages," *IEEE Trans. on Inform. Theory*, vol. 24, pp. 339–348, 1978.

[5] I. Csiszár and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, Orlando FL, 1981.

[6] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Springer-Verlag, NY, 1998.

[7] G. J. Foschini and M. J. Gans, "On limits of wireless communication in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335, 1998.

[8] J. Hagenauer, "Iterative decoding of binary block and convolutional codes," *IEEE Trans. on Inform. Theory*, vol. IT-42, no. 2, pp. 429–445, Mar. 1996.

[9] A. A. Hassan, W. E. Stark, J. E. Hershey, and S. Chennakeshu, "Cryptographic key agreement for mobile radio," *Digital Signal Processing, Academic Press*, vol. 6, pp. 207–212, 1996.

[10] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links," Technical report, Lucent Bell Laboratories Technical Memo, 2000. `http://mars.bell-labs.com/cm/ms/what/mars/index.html`.

[11] B. Hassibi and T. L. Marzetta, "Multiple-antennas and isotropically-random unitary inputs: the received signal density in closed-form," *IEEE Trans. on Inform. Theory*, vol. to appear, , 2001.

[12] A. O. Hero and T. L. Marzetta, "On computational cut-off rate for space time coding," Technical report, Bell Laboratories, Lucent Technologies, Murray Hill, NJ, 2000. `http://mars.bell-labs.com/cm/ms/what/mars/index.html`.

[13] A. O. Hero and T. L. Marzetta, "Cut-off rate and signal design for the quasi-static Rayleigh fading space-time channel," *IEEE Trans. on Inform. Theory*, vol. To appear, , 2001.

[14] J. E. Hershey, A. A. Hassan, and R. Yarlagadda, "Unconventional cryptographic keying variable management," *IEEE Trans. on Communications*, vol. 43, pp. 3–6, Jan. 1995.

[15] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," Technical report, Bell Laboratories, Lucent Technologies, Murray Hill, NJ, 1999. `http://mars.bell-labs.com/cm/ms/what/mars/index.html`.

[16] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communication," *IEEE Trans. on Inform. Theory*, vol. IT-46, pp. 543–564, Mar. 2000.

[17] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Trans. on Inform. Theory*, vol. IT-46, pp. 1962–1973, Sept. 2000.

[18] B. Hughes, "Differential space-time modulation," *IEEE Trans. on Inform. Theory*, vol. preprint, , submitted 2000.

[19] H. Koorapaty, A. A. Hassan, and S. Chennakeshu, "Secure inforamation transmission for mobile radio," *IEEE Communications Letters*, vol. 4, pp. 52–55, Feb. 2000.

[20] C. E. Landwehr and D. M. Goldschlag, "Security issues in networks with internet access," *IEEE Proceedings*, vol. 85, no. 12, pp. 2034–2051, Dec 1997.

[21] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh fading," *IEEE Trans. on Inform. Theory*, vol. IT-45, pp. 139–158, Jan. 1999.

[22] P. McCullagh, "Tensor notation and cumulants of polynomials," *Biometrika*, vol. 71, pp. 461–476, Dec 1984.

[23] P. Moulin, M. E. Mihcak, and G.-I. Lin, "An information-theoretic model for image watermarking and data hiding," in *IEEE Int. Conf. on Image Processing*, volume 3, pp. 667–670, Vancouver, B.C., 2000.

[24] P. Moulin and J. O'Sullivan, "Information thoeretic analysis of information hiding," *IEEE Trans. on Inform. Theory*, vol. Preprint, , 1999. http://www.ifp.uiuc.edu/~moulin/paper.html.

[25] L. H. Ozarow and A. D. Wyner, "The wire-tap channel, ii," *Bell Syst. Tech. Journ.*, vol. 63, pp. 2135–2157, 1984.

[26] F. Petitcolas, R. Anderson, and M. Kuhn, "Information hiding-a survey," *IEEE Proceedings*, vol. 87, no. 7, pp. 1062 –1078, July 1999.

[27] J. E. Savage, "Sequential decoding – the computation problem," *Bell Syst. Tech. Journ.*, vol. 45, pp. 149–175, Jan. 1966.

[28] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of non-minimum phase systems (channels)," *IEEE Trans. on Inform. Theory*, vol. IT-36, pp. 312–321, 1990.

[29] C. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. Journ.*, vol. 27, pp. 379–423, 1948.

[30] C. Shannon, "Communication theory of secrecy systems," *Bell Syst. Tech. Journ.*, vol. 29, pp. 656–715, 1949.

[31] A. Shokrollahi, B. Hassibi, B. B. Hochwald, and W. Sweldens, "Representation theory for high rate multiple antenna code design," Technical report, Bell Laboratories, Lucent Technologies, Murray Hill, NJ, 2000. http://mars.bell-labs.com/cm/ms/what/mars/index.html.

[32] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms*, Prentice-Hall, Engelwood, 1995.

[33] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," Technical report, AT&T Bell Laboratories Technical Memo, 1995. http://mars.bell-labs.com/cm/ms/what/mars/index.html.

[34] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, Nov.-Dec. 1999.

[35] F.-Q. Wang and D. J. Costello, "Probabilistic construction of large constraint length trellis codes for sequential decoding," *IEEE Trans. on Communications*, vol. COM-43, no. 9, pp. 2439–2448, Sept. 1995.

[36] J. M. Wozencraft and R. S. Kennedy, "Modulation and demodulation for probabilistic coding," *IEEE Trans. on Inform. Theory*, vol. IT-12, pp. 291–297, July 1966.

[37] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. Journ.*, vol. 54, pp. 1355–1387, 1975.

[38] L. Zheng and D. N. Tse, "Packing spheres in the Grassman manifold: a geometric approach to the non-coherent multi-antenna channel," *IEEE Trans. on Inform. Theory*, vol. preprint, , 2000.

Figure Captions

1. Secure Space-Time Link ($M = 3$, $N_R = N_E = 2$)

2. A link in a wireless network between a transmitter (T) and a receiver (R) who have cooperated to learn their channel H_{TR}. Eavesdropper E attempt to detect a message (known signal), detect signaling activity (known modulation), or intercept data transported by the link H_{TR} without knowing the channel H_{TE}. The eavesdropper and the client receiver must generally perform these tasks in the presence of multi-user interference. In this paper, we assume Gaussian interferers with known spatial covariances.

3. Spectra of optimal source covariance matrices under T/R-informed LPD (mean-squared-power) and mean-power constraints: SNR $= 20dB, M = N_R = 32$.

4. T/R informed power-capacity C_{pow}^{TR} $(N_R = M)$

5. T/R informed LPD-capacity C_{lpd}^{TR} $(N_R = M)$

6. Loss in LPD-capacity due to mean-squared power (LPD) constraint $(N_R = M)$

7. Loss in LPD-capacity due to mean-power constraint $(N_R = M)$

8. T/R informed power-capacity loss due to transmitter-channel mismatch $(N_R = M = 32)$.

9. T/R informed LPD-capacity loss due to transmitter-channel mismatch $(N_R = M = 32)$.